RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION

Ward Edwards

University of Southern California

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

31 August 1975

001597-F

UNIVERSITY OF SOUTHERN CALIFORNIA

social science research institute

FINAL TECHNICAL REPORT

RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION

WARD EDWARDS

CONTRACT PERIOD: 1 DECEMBER 1974 - 30 JUNE 1975

SPONSORED BY:

ADVANCED RESEARCH PROJECTS AGENCY
DEPARTMENT OF DEFENSE

MONITORED BY:

ENGINEERING PSYCHOLOGY PROGRAMS
OFFICE OF NAVAL RESEARCH
CONTRACT NO. N00014-75-C-0487, ARPA
ORDER #2105

AUGUST 1975

SSRI RESEARCH REPORT 75-

Social Science Research Institute
University of Southern California
Los Angeles, California 90007
213-746-6955

The Social Science Research Institute of the University of Southern California was founded on July 1, 1972 to permit USC scientists to bring their scientific and technological skills to bear on social and public policy problems. Its staff members include faculty and graduate students from many of the Departments and Schools of the University.

SSRI's research activities, supported in part from University funds and in part by various sponsors range from extremely basic to relatively applied. Most SSRI projects mix both kinds of goals — that is, they contribute to fundamental knowledge in the field of a social problem, and in doing so, help to cope with that problem. Typically, SSRI programs are interdisciplinary, drawing not only on its own staff but on the talents of others within the USC community. Each continuing program is composed of several projects; these change from time to time depending on staff and sponsor interest.

At present (Spring, 1975), SSRI has four programs:

Criminal justice and juvenile delinquency. Typical projects include studies of the effect of diversion on recidivism among Los Angeles area juvenile delinquents, and evaluation of the effects of decriminalization of status offenders.

Decision analysis and social program evaluation. Typical projects include study of elicitation methods for continuous probability distributions and development of an evaluation technology for California Coastal Commission decision-making.

Program for data research. A typical project is examination of small-area crime statistics for planning and evaluation of innovations in California crime prevention programs.

Models for social phenomena. Typical projects include differential-equation models of international relations transactions and models of population flows.

SSRI anticipates continuing these four programs and adding new staff and new programs from time to time. For further information, publications, etc., write or phone the Director, Professor Ward Edwards, at the address given above.

RESEARCH ON THE TECHNOLOGY OF INFERENCE AND DECISION

Final Technical Report
31 August 1975

Ward Edwards

Social Science Research Institute
University of Southern California
Los Angeles, California

This is a final report of research supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Engineering Psychology Programs, Office of Naval Research under Contract No. N00014-75-C-0487, Work Unit Number No. 197-021.

SSRI Research Report 75-10

TABLE OF CONTENTS

Introduction

The Social Science Research Institute, University of Southern California, was awarded a contract for Research on the Technology of Inference and Decision for the period December 1, 1974 to June 30, 1975 by the Advanced Research Projects Agency. The contract, N00014-75--0487, was monitored by the Engineering Psychology Programs, Office of Naval Research. Research on this topic by the Principal Investigator, Professor Ward Edwards, began at the University of Michigan, under ARPA sponsorship and continued under a previous ARPA contract at the University of Southern California when Professor Edwards left the University of Michigan to become the Director of the Social Science Research Institute. Thus many ideas that have come to fruition under this contract began their development under previous contracts, while other ideas now originating will culminate some time in the future.

This Final Report summarizes the activities conducted under this program at USC. Five technical reports have been produced and are being distributed. The abstracts of these technical reports appear at the end of this report. These technical reports are self explanatory and thus will not be dealt with in detail here. Brief descriptions of the results presented in these technical reports are included in this report to illustrate how they fit into the overall program of research. Also included are discussions of continuing activities and suggestions for future research.

A Technical Overview

The research conducted under this program falls under three major themes--each falling within a primary division of decision analysis; the two types of inputs, probabilities and utilities, and the combination of the inputs. Both theoretical and experimental work is included with much of the impetus coming from problems found in practical applications of technologies for aiding decision making.

Elicitation of Subjective Probabilities. Most of the early work on the elicitation of subjective probabilities was concerned with the probabilities of well defined single events or finite sets of mutually exclusive events. As decision analysis became more sophisticated and began dealing with more complex problems many of the needed probabilities could no longer be characterized in this manner. Instead complete probability density functions over continuous variables were often needed, e.g. the market share in a new product decision, or the cost of a new weapon system. As research began on how best to elicit these probability distributions, one result seemed to be pervasive; the elicited distributions were too tight. That is, in experiments where the true values of the uncertain quantities were known, a high percentage of the true values fell into the tails of the assessed subjective distributions. Typical results found 25% to 50% of the true values below the .01 value or above the .99 value of the assessed cumulative distributions where only 2% should be. The usual explanation for these results was that subjects overvalued their information, that is, they had a tendency to express more knowledge than they actually

had. Although training led to some improvement, the results were still discouraging.

The usual technique used to elicit these probabilities both in the experimental work and in practical applications is known as the fractile method. This technique calls for the assessor to give values of the uncertain quantity that correspond to some given fractiles of his subjective probability distribution, usually the median, upper and lower quartiles, and at least two extreme fractiles. Tversky and Kahneman (1973) suggest that in the type of judgment required for this task, a cognitive process called "anchoring and adjusting" may occur. When a subject is asked for a value corresponding to a specific fractile, the subject first anchors on the value considered most likely and then adjusts that value in the direction appropriate for the given fractile. Such adjustment processes are, however, usually insufficient leading to too tight distributions. If this cognitive process is occurring then the tightness of the assessed subjective distributions may be an artifact of the elicitation technique.

Beaver, von Winterfeldt, and Edwards (see Technical Report Abstract No. 3) conducted an experiment investigating this possibility, and obtained some rather striking results. This experiment compared the fractile procedure with a procedure in which the subjective distribution was obtained by asking questions such as "What are your odds that the true value is less than x?" where x was varied to get an approximation for the entire distribution, using almanac questions as the stimuli. For this type of question the anchoring and adjusting hypothesis suggests that for any given value of the uncertain quantity, the subject first anchors on odds of 1:1 (or probability of .50) and then adjusts the odds in the appropriate direction. In this case insufficient adjustment

will lead to too flat distributions. Also varied in this experiment were the measures of uncertainty used; probabilities, odds, and odds on a logarithmic scale,

The results showed a large difference in the tightness of distributions obtained by the two procedures, with only minor differences due to the uncertainty measure used except for the odds on a logarithmic scale, as measured by the percentage of true values falling into the tails of the subjective distributions. These "surprise" frequencies, as they are often termed, were in the range of 25% to 35% for the fractile procedures, but only 4% to 5% for the procedures requiring odds and probabilities as responses and approximately 20% for the procedure requiring odds on a logarithmic scale. Thus, although the fractile methods produced distributions that were too tight, the distributions from the second procedure were certainly not too flat. In fact for the odds and probability responses they were quite veridical.

The surprise frequencies do not show how well calibrated probability assessors are except in the tails of the distributions. Another often used measure is the percentage of true answers falling within the interquartile ranges of the assessed distributions. This measure has the advantage that it deals with the region of the uncertain quantity that is most likely to occur. In this experiment there seemed to be little difference between the various elicitation procedures used except for the odds on a logarithmic scale responses. For the other procedures the percentage of true values falling within the interquartile range varied from 42% to 57% (31% for the odds on a logarithmic scale). This suggests that in this crucial range none of the elicitation procedures do too

badly except for the odds on a logarithmic scale.

The results of the odds on a logarithmic scale procedure were quite surprising and are at this point unexplained. One possibility which we are currently investigating was suggested by some preliminary investigations for the previously described experiment. It seemed that when given a logarithmic scale of odds on which to respond, subjects simply chose the highest odds when they were very sure regardless of whether the odds were 1000:1 or 10000:1 while responding to moderate uncertainty with odds in the middle of the scale again disregarding the actual numerical value. While it is not surprising that the subjects cannot differentiate between odds of 1006:1 and 10000:1, it would be of great concern if they do not distinguish between odds of 5:1 and 50:1. We are currently investigating this hypothesis using scales with several different endpoints and both linear and logarithmic scales.

During the last ten years there has been considerable research interest in human capabilities for probabilistic inference. The major finding is that people are conservative; that is, probabilistic data cause less change in opinion than is appropriate (Edwards, 1968; Slovic and Lichtenstein, 1971). Three hypotheses have been suggested to explain this phenomenon. The misperception hypothesis asserts that people incorrectly perceive the diagnostic impact of each datum. The misaggregation hypothesis claims that single data are perceived correctly but are not combined properly with other data. The most common form of the response bias hypothesis is that people are reluctant to use the extreme odds or probabilities that are veridical as evidence accumulates.

Wheeler and Edwards (see Technical Report Abstract No. 5) conducted a series of experiments designed to test these hypotheses. In the first experiment subjects assessed both cumulative and noncumulative posterior odds and likelihood ratios. There was little difference between the odds and likelihood ratio judgments, but substantial difference between cumulative and noncumulative judgments. The cumulative responses were conservative while the noncumulative responses were near veridical. This result seems to rule out the misperception hypothesis. Experiments two and three varied the characteristics of the sequences of stimuli so the posterior odds after some sequences were still relatively small, i.e. less than 13:1. Conservatism was found even in the sequences with relatively small posterior odds, thus supporting the misaggregation hypothesis.

Recently research on the causes of conservatism has come under critical attack. It has been suggested that the phenomenon is not as pervasive as originally believed and is indeed very task- and subject-dependent. We believe that like other biases that have been discovered in probability assessment, it is clearly a task-dependent finding. This is, however, a fruitful topic of research since it is necessary to discover these biases and their causes in order to deal with them in practical applications.

Multi-attribute Utility Theory. As utility theory has progressed both in its theoretical development and in its applications, the gap has widened between the theoreticians and actual users. Users are often not concerned with how utility measurement was developed; only with how it can be applied. In practice this can lead to misuse. A theoretically inappropriate model or assessment procedure may be used possibly leading ultimately to an incorrect

decision. Users need to be aware of the implications of various models and assessment procedures and understand which are good approximations and which are not. Fortunately decision analysis is relatively insensitive to such errors. However, it is still advantageous for the user to understand the relationship among the various models and assessment procedures.

The review by von Winterfeldt (see Technical Report Abstract No. 4) serves this purpose. Existing utility models are classified according to their underlying measurement theoretic representations. The assumptions of the models, both behavioral and technical, are discussed at a level not requiring familiarity with measurement theory. Another valuable product of this report is the discussions of logical relationships and similarities among models and assessment procedures. These similarities allow users of utility theory to approximate complex models and assessment procedures with much simpler ones.

The most severe problem facing developers of multi-attribute utility (MAU) procedures is the lack of a completely satisfactory method of validation. Typically validation has taken the form of measures of convergence (usually correlations) with various other procedures, e.g. "wholistic judgments," purporting to assess the same underlying quantity. Having concluded that such methods are not entirely appropriate, we have searched for other validation procedures. The best-of-all-possible-worlds would be to have a true criterion against which to compare utilities assessed by various MAU techniques. An extreme subjectivist would argue that such an external standard cannot exist because utilities are inherently internal to the individual. We do not completely agree with this argument. However, from an experimental validation point of view this philosophical disagreement appears to be pointless. We

have yet to find a situation that meets the requirements for such a validation where an external criterion exists. Given the current infeasibility of this approach, what other validation procedures can be explored?

Another area of psychology, the theory of mental tests (see, for example, Ghiselli, 1964; or Gulliksen, 1950), has long dealt with a similar problem; how well does a combination of subtests (or items) measure an ability for which no "true" criterion exists. In addition, the usual procedure for combining the subtests or items into a single score is to take a weighted average of the individual subtests or items. This model is formally identical to the most prominent MAU model, the additive model. Because of these similarities, we feel exploring the approaches traditionally used by mental test theorists may be enlightening in our search for answers to this pressing problem.

In this spirit Newman (see Technical Report Abstract No. 2) has investigated a theory and set of procedures for assessing the dependability of MAU procedures. (We use the word dependability to represent both validity and reliability.) It is called the Theory of Generalizability and has been developed by Professor Lee Cronbach and his students at Stanford University. The theory abandons the concept of a "true score", eliminates the need for restrictive assumptions such as "parallel measures," and does not require the investigator to define a criterion of success to be used in validity studies. The theory replaced the concept of a true score with that of a universe score. To ask the question of how reliable or valid a measure is, is to ask how well one can generalize from the observations at hand to the universe or domain of observations to which they belong. To ask about the agreement of judges in MAU studies is to

ask how well we can generalize from one set of judgments to judgments from all possible judges who might have been chosen for the particular study. The theory requires the investigator to specify the universe of conditions of observation over which he wishes to generalize. Conditions is a generic term referring to observers (judges), forms of stimuli, occasions, etc.. In addition to generalizing to a universe of judges for example, we may also wish to generalize to a universe of situations in which the judgments were made. Miller, Kaplan, and Edwards (1968) studied the efficacy of a utility model in four tactical military logistic situations. It is of interest to know how well one could generalize to all possible tactical situations which the four represented. Gardiner (1974) used 15 typical housing development permit requests in his application of MAU techniques to coastal zone management decision making, and again it is desirable to know the degree of generalizability to the universe of all such permit requests.

The theory uses analysis-of-variance models and relies heavily on estimates of variance components using expected values of the mean squares yielded by these models. The only assumption made is that the conditions of the study are randomly sampled from a universe of conditions. Using the estimates of variance components, it is possible to define a coefficient of generalizability that indicates how well one can generalize from the observed data to the universe score. The familiar distinction between reliability and validity along with separate estimates of reliability and validity coefficients is eliminated. The definitions of reliability and validity coalesce and only one coefficient--the coefficient of generalizability needs to be estimated in any study.

We have applied this theory to already completed studies and in each case MAU techniques have been demonstrated to have higher coefficients of generalizability than other techniques designed to do the same thing. In the Miller, Kaplan, and Edwards (1968) study for example, a subjective value-judgment based Tactical Air Command System has a higher coefficient of generalizability than the conventional system at least as demonstrated by laboratory studies. Also Gardiner's (1974) utilization of MAU procedures in coastal zone management decisions was found to have a higher coefficient of generalizability than so called "wholistic judgment" procedures. It should be pointed out that the conventional systems also had coefficients of generalizability which could be considered respectable, but the MAU procedures had higher coefficients and therefore, in our opinion, were more dependable.

We intend to explore this theory in more detail. Next on the agenda is to develop ways of establishing credible interval estimates for the coefficient of generalizability either by assuming a theoretical distribution for the coefficient or by obtaining empirical estimates for the coefficient by doing cross validation studies using Tukey's "Jack Knife" method, or a combination of both.

Another validation approach arising from the extreme subjectivist's position determines a MAU model from the behavioral properties that characterize the decision maker's evaluation strategy. Given that certain sets of behavioral assumptions are true, representation theorems from measurement theory show that utilities exist with certain formal properties. Models and their appropriate assessment procedures can be arranged hierarchically according to strength.

By adding assumptions, weaker models become stronger. Adding assumptions, of course, increases the likelihood that some assumption will be violated. We, therefore, find a natural tradeoff between model strength and probability of violations of model assumptions. Since stronger models are preferred in use primarily due to the simplicity of parameter assessment, the question arises as to which assumptions will often be violated and which models should and should not be used.

We are currently experimentally investigating some of these and other similar questions. The main thrust of this validation idea is to experimentally determine behaviorally meaningful properties which characterize the decision maker's evaluation strategy and, therefore, should be implemented in a model of his evaluation process. For example, are decision makers induced to take more (or less) risk when evaluating gambles of single commodities when they are given a bonus in the form of a certain amount of another commodity? Do decision makers judge gambles with multi-attribute outcomes solely on the basis of probabilities and amounts in single attributes, or are they also sensitive to the amount of outcome variation? By determining such orderly, intended, and consistent properties of the decision maker's evaluation strategy in risky multi-attribute evaluations, we will be able to eliminate all those evaluation models which cannot account for these properties.

Error in Decision Analysis. Decision analysis, like any other modeling process, can be wrong as a basis for action in either or both of two ways: the model may be wrong, in the sense of being either misleading or too crude a representation of the phenomenon modeled; or the data may be wrong, in any of a variety of ways. The literature shows some considerations of the latter of

these possibilities, but the former is almost never discussed. Our current work shows decision analysis to be much more sensitive to errors in modeling, than to data errors, i.e. inappropriate utilities and/or probabilities. Although we must be careful to eliminate gross errors in the assessment of probabilities and utilities (the technical report by Seaver, von Winterfeldt, and Edwards shows that such errors can and do exist.), large deviations from optimal decision strategies or model parameters will lead to relatively small losses in expected value given some "relatively mild" assumptions (von Winterfeldt and Edwards, 1973). In light of this fact we were surprised by Fryback's (1974) finding that in a real world medical decision problem, although the functions showing the relations between size of error in decision strategy and resulting loss in expected utility were quite flat, the doctors were actually obtaining only a little more than 50% of the expected utility obtainable by the decision-theoretically optimal procedure.

On reflection, we realized that our flat-maximum analysis had failed to deal with two important facts. One is that real decision are typically made without proper prior decision-analytic structuring, and in particular without prior elimination of grossly inappropriate decisions or strategies. The other is that the flat maximum ideas apply only to the decision making part of a decision analysis, not to the information processing part. Neglect or inefficient use of information can in effect create dominated strategies, not recognizable as such from inspection of payoff matrices or decision trees, and can make these dominated strategies seem optimal.

von Winterfeldt and Edwards (see Technical Report Abstract No. 1) have

examined this new idea showing how "inefficient information" can lead to dominated strategies for three specific definitions of inefficient information. Although the results are proven for these specific examples, the generalization is clear. The thrust of this idea is similar to and somewhat expanded from the point made by previous work on the flat maximum: in decision analysis, structuring the problem and processing the information are of primary importance, while eliciting probabilities and utilities and deciding among admissible alternatives are of secondary importance. Any broad research effort in decision analysis should recognize these priorities. Research on the merits of information sources, on optimization of information processing, and on formulation of decision problems is more important than work on precise elicitation and optimization procedures.

References

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.) Formal Representation of Human Judgment. New York: Wiley, 1968, 17-52.

Fryback, D. G. Use of radiologists' subjective probability estimates in a medical decision making problem. Michigan Mathematical Psychology Program, MMPP 74-14, University of Michigan, Ann Arbor, 1974.

Gardiner, P. C. Public policy decision making: The application of decision technology and Monte Carlo simulation to multiple objective decisions- A case study in California Coastal Zone management. Unpublished doctoral dissertation, University of Southern California, 1974.

Ghiselli, E. E. Theory of Psychological Measurement. New York: McGraw Hill, 1964.

Gulliksen, H. Theory of Mental Tests. New York: Wiley, 1950.

Miller, L., Kaplan, R., and Edwards, W. JUDGE: A laboratory evaluation. The RAND Corporation, RM-5547-PR, 1968.

Slovic, P. and Lichtenstein S. A comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. Oregon Research Institute Bulletin, 1973, Volume 13, Number 1.

von Winterfeldt, D. and Edwards, W. Flat maxima in linear optimization models. Technical Report No. 011313-1-T, Engineering Psychology Laboratory, University of Michigan, Ann Arbor, 1973.

Report Abstract 1

Error in Decision Analysis: How to Create The Possibility of Large Losses by Using Dominated Strategies

Detlof von Winterfeldt and Ward Edwards
University of Southern California

This report examines some concepts, sources, and possible consequences of error in decision analysis. Recent articles on the possibilities for error in decision analysis showed that under some relatively mild assumptions deviations from optimal decision strategies or from optimal model parameters will lead only to minor losses in expected value. This "flat maximum" property of decision analytic models applies, however, only to admissible decisions. By inadvertently selecting a dominated (inadmissible) decision, the decision maker creates the possibility for large expected losses. Usually dominance can be recognized and losses can be avoided by elimination of dominated decisions. Unfortunately, for a large class of errors the discovery of dominance is difficult if not impossible. These errors consist of failing to use information or using it inappropriately in decision strategies. The main point this report makes is that such errors can, and typically will, lead to dominated strategies, and so can lead to substantial expected losses.

Report Abstract 2

Assessing the Reliability and Validity of Multi-attribute Utility Procedures: An Application of the Theory of Generalizability

J. Robert Newman
University of Southern California

This report presents a theoretical rationale for assessing the reliability, validity, and dependability of multi-attribute utility models and techniques. If an investigator is advocating the use of a MAU model or procedure he or she is interested in generalizing from observations at hand to a universe or domain of observations that are members of that same universe. The universe must be unambiguously defined but it is not necessary to assume that universe as having any statistical properties such as uniform variances or covariances. A study of generalizability is conducted by taking measurements on persons, stimuli, tasks, etc. that are assumed to be randomly representative of a universe an investigator wishes to generalize to. The ratio of an estimate of the universe "score" variance to an estimate of the observed score variance is the coefficient of generalizability. This is estimated by the intra-class correlation coefficient. ANOVA and the Expected Mean Square paradigm of Cornfield and Tukey is used to obtain the appropriate variance estimates.

The theory dispenses with unnecessary and unwarranted assumptions, and eliminates the distinction between reliability and validity. Any generalizability study can be conducted without reference to having a parallel measure of the MAU instrument or some external criterion of "success". If a MAU

technique is compared to some non-MAU technique for doing the same thing then it is possible to calculate the coefficient of generalizability for both methods thus allowing the investigator to decide which is best for his or her purposes. Three numerical examples are given of the theory. Preliminary investigations have indicated that MAU models and techniques based on such models may be "better" than non-MAU models since the former have a tendency to reduce the interaction between judges and the thing being judged when such interaction represents inconsistency of judgment.

Report Abstract 3

Eliciting Subjective Probability Distributions on Continuous Variables

David A. Seaver, Detlof v. Winterfeldt, and Ward Edwards
University of Southern California

Five procedures for assessing subjective probability distributions over continuous variables were compared using almanac questions as stimuli. The procedures varied on the uncertainty measures used (probabilities, odds, and odds on a logarithmic scale) and the type of response required from the subjects (uncertainty measure or value of the unknown quantity). The results showed the often used fractile procedures were inferior to procedures requiring probabilities or odds as the response from subjects. The results are also discussed in terms of the "anchoring and adjustment" hypothesis.

Report Abstract 4

An Overview, Integration, and Evaluation of Utility Theory for Decision Analysis

Detlof von Winterfeldt

This report is a survey of the measurement theoretic literature on utility models and assessment. It was specifically written for decision analysts who are interested in the use of these abstract models and methods for evaluation problems in real world decision making. The report is, first, an inventory and dictionary that classifies, translates, and integrates existing measurement theories; and second, an evaluation of the usefulness of measurement theory as a tool for solving complex decision problems. The first part of the report classifies and describes utility models. After discussing some general aspects of utility theory as part of measurement theory, a classification scheme for utility models is developed with emphasis on the characteristics of the decision problem to which the model applies. Then the main utility representations --weak order measurement, difference measurement, bisymmtric measurement, conjoint measurement, and expected utility measurement--are described through their assumptions, model forms, formally justified assessment procedures, and common approximation methods. The second part of the report discusses some similarities and differences among these models and assessment procedures. Topics include logical relationships between models, similarities in the cognitive processes involved in different assessment procedures, and model convergence by insensitivity. The third and final part of the report evaluates the use of utility theory for decision analysis, as a tool in formal treatments of decision problems.

Utility theory can be quite useful in structuring evaluation problems and in eliciting appropriate model forms, but the theoretically feasible assessment procedures are often too clumsy and complicated to be applicable in real world preference assessment. A general critique of current trends in utility theory concludes the report.

Report Abstract 5

Misaggregation Explains Conservative Inference About Normally Distributed Populations

Gloria E. Wheeler and Ward Edwards

University of Southern California

Three major hypotheses have been proposed to account for conservative inference: misaggregation, misperception, and response bias. The research reported in this paper allowed the testing of these hypotheses. Subjects made probabilistic judgments about stimuli generated from normally distributed populations. The populations were piles of pick-up sticks, each stick having one end painted blue and the reminder painted yellow. The length of blue paint was the random variable. In Experiment 1, each S made 4 types of judgments: noncumulative likelihood ratios, noncumulative odds, cumulative likelihood ratios, and cumulative odds. The results indicated that there was little difference between likelihood ratio and odds judgments, and that when judging single stimuli, Ss were veridical; conservatism only occurred when Ss were in a cumulating condition. Thus the results ruled out the misperception hypothesis.

Experiments 2 and 3 varied d', sequence construction, and population display. Sequences were constructed that would accentuate differences between predictions made by response bias and misaggregation hypotheses. The data showed that subjects made veridical independent trial estimates but aggregated information conservatively, regardless of how far odds and likelihood ratios were from 1:1, thus permitting rejection of most forms of the response bias hypothesis.

Social Science Research Institute
Research Reports

75-1 N. Miller and G. Maruyama. Ordinal Position and Peer Popularity. January, 1975.

75-2 G. Maruyama and N. Miller. Physical Attractiveness and Classroom Acceptance. January, 1975.

75-3 J.R. Newman, S. Kirby and A.W. McEachern. Drinking Drivers and Their Traffic Records. February, 1975.

75-4 Detlof von Winterfeldt and Ward Edwards. Error in Decision Analysis: How to Create the Possibility of Large Losses by Using Dominated Strategies. April, 1975.

75-5 Peter C. Gardiner and Ward Edwards. Public Values: Multi-Attribute Utility Measurement for Social Decision Making. (Forthcoming as a chapter in *Human Judgment and Decision Processes: Formal and Mathematical Approaches,* Steven Schwartz and Martin Kaplan (eds.), summer 1975.) May, 1975.

75-6 J. Buell, H. Kagiwada, and R. Kalaba. SID: A Fortran Program for System Identification. May, 1975.

75-7 J. Robert Newman. Assessing the Reliability and Validity of Multi-Attribute Utility Procedures: An Application of the Theory of Generalizability. July, 1975.

75-8 David A. Seaver, Detlof v. Winterfeldt, and Ward Edwards. Eliciting Subjective Probability Distributions on Continuous Variables. August, 1975.

75-9 Detlof von Winterfeldt. An Overview, Integration, and Evaluation of Utility Theory for Decision Analysis. August, 1975.

75-10 Ward Edwards. Research on the Technology of Inference and Decision. August, 1975.